

Related works

Biomedical Extractive Question Answering (BEQA) has gained attention for accurate and efficient extraction of information from complex biomedical texts. Research has explored integrating Named Entity Recognition and Relation Extraction tasks, transformer-based models, and hybrid learning approaches combining deep learning techniques like BiLSTM and ensemble learning. These approaches have improved the handling of domain-specific terminology and enhanced BEQA system performance.

BioMedical Question Answering

A framework for answering questions with perfect and exact answers has been provided in a number of recent research, and state-of-the-art results on questions with perfect answers on the BioASQ dataset have been obtained at ([Sharma et al., 2018](#)). They use LeToR ranking models to respond to factual or list-based queries, neural entailment models for precise replies, and a unique embedding transformation method for yes/no questions. They enhance the IR part of extractive summarization for optimal responses. The generated summary answer's human readability still needs a great deal of improvement, even if this greatly raises ROUGE scores.

For yes/no, factoid, and list questions at ([Sarrouiti & Alaoui, 2020](#)), SemBioNLQA can provide specific answers (e.g., "yes," "no," a biomedical entity, etc.) as well as paragraph-sized ideal replies (summaries of pertinent information). On the other hand, it only returns optimal responses to summary queries.

Meanwhile, the framework LIQUID—which automatically creates a list of QA datasets from unlabeled corpora—was presented to address the issue of data scarcity in this field ([Lee et al., 2023](#)). On five benchmark datasets, the synthetic data significantly enhanced the performance of the existing supervised models. Here, the impact of every LIQUID component was carefully examined, and quantitative and qualitative data were produced.

BioMedical QA with Name Entity Recognition

Biomedical Question Answering (QA) systems aim to extract relevant information from vast amounts of biomedical literature in response to user queries. Named Entity Recognition plays a crucial role in this process by identifying and classifying key entities within the text, such as genes, proteins, diseases, and drugs. Accurate NER serves as a foundation for downstream tasks like relation extraction and ultimately improves the overall performance of BioMedical QA systems.

Several recent studies have explored the integration of NER techniques to enhance BioMedical QA. ([Agirre et al., 2019](#)) investigated the impact of NER on a question answering system focused on pharmacological substances and compounds. Their findings demonstrated that incorporating a state-of-the-art NER model significantly improved the system's ability to answer questions involving these entities.

Deep learning architectures have developed robust NER models for biomedical literature. ([Wei et al., 2019](#)) proposed a Bidirectional Long Short-Term Memory (BiLSTM) network for better detection of biomedical entities, ([Zhang & Chen, 2022](#)) used multi-task learning and pre-trained language models for excellent accuracy. Their model effectively captured contextual information from biomedical text and achieved high accuracy in recognizing diverse biomedical entities.

These studies highlight the significant contribution of NER to BioMedical QA. By pinpointing relevant entities within the text, NER systems enable QA systems to focus on crucial information and provide more accurate and informative answers to user queries.

Relation Extraction in BioMedical QA

BioMedical Question Answering (QA) systems not only need to identify relevant entities within the text, but also comprehend the connections between these things to provide comprehensive and informative answers. Relation extraction (RE) plays a vital role in this process by uncovering the semantic links between entities, such as protein-protein interactions, drug-disease associations, and gene regulatory pathways. By capturing these relationships, RE empowers BioMedical QA systems to reason over the extracted information and deliver more insightful responses to complex user queries.

Recent research has highlighted the importance of RE for enhancing BioMedical QA performance. ([Sousa, 2021](#)) proposed a deep learning-based approach for RE that utilizes convolutional neural networks (CNNs) to capture local patterns and recurrent neural networks (RNNs) to model long-range dependencies within sentences. Their system achieved promising results in extracting various biomedical relations from text.

Further advancements have been made in exploiting external knowledge resources to improve RE accuracy. ([Sousa & Couto, 2022](#)) introduced a method that integrates knowledge graphs with a neural RE model. This approach leverages pre-existing knowledge about biomedical entities and relationships in order to improve the model's performance to extract novel relations from unseen text.

These studies demonstrate the significant influence of RE on BioMedical QA performance. By extracting relationships between identified entities, RE empowers QA systems to move beyond simple entity recognition and provide a deeper understanding of the underlying biological processes. This capability is crucial for answering complex biomedical queries that require reasoning over interconnected entities and relationships.

Transformer and BiLSTM Architectures for BioMedical QA

This work explores the effectiveness of Transformer and Bi-directional Long Short-Term Memory (BiLSTM) architectures in Question Answering (QA) tasks. It compares and contrasts the ability of these deep learning models to convey contextual relationships and long-range dependencies in text for improved answer extraction.

In the context of the MEDIQA challenge, the LasigeBioTM team proposed an approach that explored a common Transformer-based architecture for each task, utilizing the given training data to adjust the same pre-trained weights for every activity at ([Lamurias & Couto, 2019](#)). The group used NER, a named entity recognition tool, to enhance the question and answer texts and supplemented the training data with additional datasets.

In the context of the BioASQ10B challenge, Alrowili and Vijay-Shanker (2022) explored the use of BioM-Transformers models, an adaptation of both ELECTRA and ALBERT models to the biomedical domain at ([Alrowili & Vijay-Shanker, 2022](#)). The authors extended their investigation of biomedical QA models with BioM-Transformers by combining the BioASQ10B-Factoid training set with the List training set to overcome its small size, and by performing a grid search for hyperparameters.

Several studies have explored the application of BiLSTM-CRF models for NER in the context of online health communities. For instance, Wang et al. [2020] proposed a BiLSTM-CRF model for NER in a diabetes-focused online community, achieving promising results in identifying medical entities ([Wang & Guan, 2020](#)). Their work

highlights the potential of this approach for extracting valuable medical information from doctor-patient interactions. Bidirectional Long Short-Term Memory (BiLSTM) networks are a popular choice because they can identify long-range dependencies within text data. Conditional Random Fields (CRFs) are often employed in conjunction with BiLSTMs to model the sequential relationships between entities and improve recognition accuracy.

Biomedical QA with Hybrid Learning Approach

([Yang et al., 2023](#)) investigates Utilizing Transfer Learning and Ensemble Learning techniques to improve the performance of Question Answering (QA) systems. The work examines how combining multiple diverse QA models (ensemble learning) can lead to more robust and accurate answer retrieval.

In the realm of biomedical question answering with Transformer ensembles, several notable works have contributed to advancing the field. Zhang et al. (2023) introduced an ensemble learning pipeline that leverages Large Language Models (LLMs) to improve medical question-answering tasks' precision and dependability ([Yang et al., 2023](#)). Their study focused on improving performance across diverse medical QA datasets, including PubMedQA, MedQA-USMLE, and MedMCQA, by employing two primary ensemble methods: Boosting-based weighted majority vote ensemble and Cluster-based Dynamic Model Selection. The results demonstrated superior performance compared to individual LLMs, showcasing accuracies of 35.84%, 96.21%, and 37.26% for MedMCQA, PubMedQA, and MedQA-USMLE, respectively, with the Majority Weighted Vote method.

A transfer learning-based sentiment-aware model named SentiMedQAer for biomedical question answering (QA) ([Zhu et al., 2022](#)), aiming to address limitations in existing biomedical QA datasets, particularly their small size and the prevalence of factoid questions. The study aims to enhance the accuracy and reliability of medical question-answering (QA) tasks by developing an ensemble learning pipeline that utilizes state-of-the-art large language models ([Yang et al., 2023](#)). The goal is to improve performance on diverse medical QA datasets.

([Kim et al., 2023](#)) explored various techniques for data augmentation, ensemble learning, model training, and pre-processing for biomedical QA, using advanced pre-trained models like GPT-4 and BioLinkBERT. Their study achieved top rankings in each of the four batches of BioASQ Task 11b-Phase B questions that are yes/no in nature, one of the four batches that are factoid in nature, and two of the four batches that are list-type questions.

Another approach to transfer learning for biomedical factoid question answering is named entity aware transfer learning. This approach involves using named entities, such as genes, drugs, and diseases, to guide the transfer learning process. For example, ([Peng et al., 2022](#)) proposed a named entity aware transfer learning method that uses a pre-trained language model to extract features from biomedical literature, with a focus on named entities. The authors reported that their approach achieved significant improvements over traditional transfer learning methods.

Ref No	QA Form	Research	Model Used	Dataset	Findings	Limitations/Research Gap
3	Biomedical QA	BioAMA: End to End Biomedical QA System	IR-based, NLI-based using Hierarchical CNN, Two-stage using NER taggers and ranking algorithms	BioASQ 5b	ROUGE-2: 0.72, ROUGE-SU4: 0.71, NLI framework accuracy: 65.6%	Inadequate training data for NLI model, needs embedding projection technique, continued refinement and optimization for different question types
5	List QA	LIQUID: Framework for List QA Dataset Generation	Summarization Model: BARTbase, QG Model: Trained on SQuAD, QA Model: Trained on SQuAD, NER Models	MultiSpanQA, Quoref, BioASQ 7b, 8b, 9b	Improved F1 scores: MultiSpanQA: 5.0, Quoref: 1.9, BioASQ datasets: 2.8	Efficiency issues, quality of synthetic data, need optimization for iterative filtering and answer expansion stages
6	Named Entity Recognition	PharmaCoNER: Pharmacological Substances, Compounds, and Proteins Named Entity Recognition Track	Summarization Model: BART, NER Model, QG Model: Trained on SQuAD, QA Model: Trained on SQuAD	MultiSpanQA, Quoref, BioASQ 7b, 8b, 9b	Improved performance with synthetic data, enhanced quality of generated QA pairs	Reliance on high-quality summarization and NER models, iterative filtering process needs optimization
7	Named Entity Recognition	Fusion Attention-Based BiLSTM-CRF for Biomedical Named Entity Recognition	BiLSTM, Attention Mechanism, CRF	JNLPBA, BC2GM	F1-score of 73.50 on JNLPBA, attention mechanism prevents significant information loss	Need for validation on other biomedical corpora, improvement in automatic feature learning, optimization of attention mechanism

8	Named Entity Recognition	BioNER with Combined Feature Attention and Fully-shared Multi-task Learning	Vanilla BioBERT, BioKMNER, BioBERT-TWA, BioBERT-CFA, Fully-shared MTL	BC2GM, JNLPBA, BC5CDR-Disease, NCBI-Disease, Linnaeus, Species-800, BC5CDR-Chemical	BioBERT-CFA model achieved competitive performance, MTL model improved BioNER tasks by leveraging cross-type information	Further incorporation of syntactic features, need for generalization across different biomedical entities
10	Relation Extraction	Biomedical Relation Extraction with Knowledge Graph-Based Recommendations	Ontology Embedding Layer, LSTM, TUP Model	BC2GM, JNLPBA, BC5CDR-Disease, NCBI-Disease, Linnaeus, Species-800, BC5CDR-Chemical	Improved performance with single-task and multi-task models, better performance with ensemble methods	Impact of different syntactic information on multi-task model performance, integration of more diverse biomedical data
11	Biomedical QA	LasigeBioTM at MEDIQA 2019: Biomedical QA using Bidirectional Transformers and NER	BERT, BioBERT	NLI training set, RQE training set, QA training set, MedQuAD Cancer-Gov dataset	High accuracy on NLI task, improved RQE and QA performance with NER augmentation	Drop in accuracy from development to test set, ranking issues in detecting correct answers
12	Biomedical QA	Exploring Biomedical QA with BioM-Transformers at BioASQ10B Challenge	BioM-Transformers: Adaptations of ELECTRA and ALBERT	BioASQ10B training and test datasets, BioASQ classification dataset	Improved performance by combining training sets, extensive hyperparameter tuning, varied performance across batches	Technical and time constraints, reproducibility concerns with Transformer-based models

27	Medical QA	One LLM is not Enough: Harnessing the Power of Ensemble Learning for Medical QA	Boosting-based Weighted Majority Vote Ensemble, Cluster-based Dynamic Model Selection	MedMCQA, PubMedQA, MedQA-USMLE	Superior performance with ensemble methods, significant accuracy improvement on medical QA datasets	Challenges in ensembling LLMs, cost associated with training large models, need for further exploration in structured QA tasks
15	Biomedical QA	Exploring Approaches to Answer Biomedical Questions: From Pre-processing to GPT-4	BioLinkBERT, GPT-4	BioASQ-10b, BioASQ-11b, SQuAD, LIQUID	High rank in BioASQ Task 11b, BioLinkBERT outperformed existing models, GPT-4 effective in list-type questions	Limited exploration of data augmentation for yes/no questions, further exploration of GPT models for different question types
16	Factoid QA	Named Entity Aware Transfer Learning for Biomedical Factoid QA	BioBERT, BiLSTM, Ensemble Method (Bagging)	BioASQ 6b, 7b	Improved performance with fine-tuned BioBERT and BiLSTM, effective ensemble approach	Complexity of biomedical named entities, data imbalance, challenges in learning sentence embeddings
28	Biomedical QA	Improving Biomedical QA with Sentence-based Ranking at BioASQ-11b	Sentence embedding cosine similarity ranking, BioM-ELECTRA model	BioASQ 2023 11b dataset, BioASQ 7b, PubMedQA	Improved accuracy for yes/no and factoid questions, sentence ranking step enhanced QA system performance	Limited improvement in ideal answer generation performance, dependency on extractive QA approach

29	Relation Extraction	Biomedical Relation Extraction with Entity Type Markers and Relation-specific QA	Span QA-based Relation Extraction, Binary QA-based Relation Extraction	DrugProt dataset	Better performance with Binary QA-based method, entity type markers improved model effectiveness	Focused on DrugProt dataset, scalability issues with manual creation of question templates
30	Sentiment-aware QA	Transfer Learning-based Sentiment-aware Model for Biomedical QA	BioBERT, T5, RoBERTa, XGBoost	PubMedQA dataset	Outperformed SOTA by 15.83%, integration of sentiment information significantly enhanced performance	Generalizability limited to PubMedQA dataset, further exploration needed for other question types

Table: Related works in Biomedical QA, NER, and RE

Our research uniquely combines multiple advanced techniques to address key limitations in existing biomedical question answering systems. Unlike previous approaches that rely solely on single models or limited combinations, we integrate NER, RE, and BiLSTM layers with the Bio+ClinicalBERT model. This comprehensive approach mitigates the need for embedding projection and optimization for different question types. By incorporating NER and RE, we enhance the model's ability to handle domain-specific terminology and complex relationships, addressing limitations. Our use of ensemble learning with bootstrap sampling and majority voting improves upon the single-model approaches, leading to more robust performance across diverse biomedical questions. Furthermore, our method achieves high accuracy without relying on extensive hyperparameter tuning or large-scale datasets, which were limitations. By addressing these gaps, our research advances the field of biomedical extractive question answering, offering improved accuracy and reliability in processing complex biomedical information.